DOCUMENT RESUME

ED 247 952 IR 050 835

AUTHOR Chadwyck-Healey, Charles
TITLE Databases in the United Kingdom.
PUB DATE 23 Jun 84
NOTE 18p.; Paper presented at the Annual Conference of the American Library Association (Dallas, TX, June 23, 1984).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attitude Measures; *Databases; Government Publications; *Information Storage; Librarians; Library Associations; Library Surveys; *Online Systems; *Use Studies
IDENTIFIERS Machine Readable Data; *United Kingdom

ABSTRACT

This overview of the status of online databases in the United Kingdom describes online users' attitudes and practices in light of two surveys conducted in the past two years. The Online Information Centre at ASLIB sampled 325 users, and Chadwyck-Healey, Ltd., conducted a face-to-face survey of librarians in a broad cross-section of 76 libraries. It was found that each of an estimated 1,000-1,500 users of online databases in the United Kingdom performs less than 150 searches per year, and their primary areas of interest are engineering/science, multidisciplinary, social sciences, agriculture, and biomedical sciences. Databases produced in the United Kingdom across a wide spectrum of subject areas are discussed, including POLIS (an index of Parliamentary information compiled by the staff of the House of Commons); HANSARD (a full-text database of a daily verbatim record, published by Her Majesty's Stationery Office, of what is said in debates in the House of Commons); DHSS Data (covers health services, social welfare, and social issues); and World Reporter (a news and current affairs information service). BLAISE, the British Library's online service that supplies the machine-readable files for catalogers throughout the United Kingdom and abroad, is also discussed. (THC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED247952

VENUE: ALA Conference, Dallas, Texas

DATE: June 23 1984

PROGRAM: European online: New Sources from the Old World

TITLE: Databases in the United Kingdom

Charles Chadwyck-Healey

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C.E. Chadwyck-Healey

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Ladies and Gentlemen, Good Morning!

It is pleasant to be the first speaker on the first day of a week-long marathon like ALA.

Not so pleasant to have to speak on a subject on which I do not profess to be an expert. But then the late Herman Kahn, the great futurologist, recognized the limitations of experts when, in 1967, he organized a conference at the Hudson Institute on the political and economic development of the world until the year 2000. For 2 days the audience listened to experts and pronouncing on their specialised subjects. On the last day Kahn said, 'You have heard what the experts think the future will be like. Now I am going to show you what the future will really be like'. He then introduced 3 surprise speakers. The first was a white revolutionary, preaching death to capitalism, the second was a black revolutionary, telling of cities burning and blood flowing in the streets, and the third was a hippie with flowers in his hair describing the peace and brotherhood that would follow from the universal use of LSD.

They left an immensely powerful impression, making the audience aware of the gulf between academic analysis and the real world - always a danger when someone with his own narrow viewpoint - in my case that of a publisher - is asked to give a broad overview on any subject. So, in the case of databases in the United Kingdom, whose viewpoint is the most valid? Surely, it has to be that of the user, to whom the publishers and the database vendors must eventually concur. So I am going to start by giving you an insight into UK online-users' attitudes and practices described in two surveys carried out over the last 2 years.

IR050835

The first survey of UK online users was carried out in 1982 by Jacky Deunnette and Sharon Hall of the Online Information Centre at ASLIB, one of the two main library associations in Britain. The Centre is also funded by the British Library and the Department of Trade and Industry.

The second is our own survey carried out in 1983 of a broad cross-section of libraries in Britain.

I am also going to make occasional references for comparison to a 1979 survey by Jacky Deunnette.

While the typical user, his habits and his preferences are not very different from what they were in 1979 there is an enormous difference in the range of vendors and databases available to him. For example online users in the UK have access to over 50 different host systems compared to 8 in 1979.

On the other hand many users complain that they do not have the time or resources (and I would add, the wish) to experiment as much as they would like to - this is very much born out by the startling preference for certain hosts which shows up in our survey as well.

I must put the sample of 325 respondents into perspective. There are probably 1000-1500 users of online databases in the UK so 325 is about one quarter, a significant sample.

The typical user in 1982 is very similar to that in 1979. It is most likely to be an industrial or commercial organisation sited in the South of England. Subject interests are mainly scientific or technical probably either biochemical or chemical or engineering. Databases searched are large well-established ones such as Chemical Abstracts and the host used is probably Dialog or one or two others.

BEST COPY AVAILABLE

The typical user performs under 150 enquiries per year and uses under 50 hours of online time. It is unlikely to pass on charges to end users and training is a combination of courses run by vendors and publishers and in-house training.

Now I can almost sense your disappointment. It is all rather predictable - to find for example that the hosts and databases used are much the same that you use, but, when I tell you about some of the home-grown British databases that are available you will see that the British online scene does have its own national flavour.

Let's look further at some of the survey statistics. Types of organisation: 40% Industrial Commercial; 25% Academic; Public Libraries, surprisingly, less than 10%.

Purpose of online searches:

Not surprisingly 99% of users carried out literature searches; only 38% fact retrieval and under 5% data analysis.

Subject areas of most interest: yes , we know that Engineering/Science will be predominant with 104 of the respondents. Multidisciplinary and Social Sciences come next, followed by Agriculture and Biomedical Sciences. Law is surprisingly low but I suspect that has more to do with a possible absence of law libraries private or institutional amongst the respondents.

Which host systems are used? Dialog comes out on top with a remarkable 87% compared to 79% in 1979, so clearly has more than held its own during a period when the number of rival hosts has multiplied more than 6 times. It is also remarkable that Dialog is so far ahead of BLAISE the British Library's own bibliographic database which carries 16 databases including the US and UK MARC files, but BLAISE only comes second with 66% of users having access to it.

With ESA-IRS the European Space Agency Service carrying 81 databases close behind with 61%. Then come SDC, Pergamon Infoline and Datastar in that order ranging from 51% to 25%; the rest are also-rans, none of which reached more than 13% of the respondents.

A further point about Dialog's popularity. It would be easier to explain if the majority of users only had access to one host and that host was Dialog but this in fact only applied to a mere 10%, meaning that 90% had the choice of more than one host.

What databases did they search on these hosts? Chemical Abstracts was mentioned by over 40%; Biosis, Inspec, Medline, Compendex by over 20%.

Under 20% of users mentioned a significantly wider selection of databases used and these included ERIC, Excerpta Medica, NTIS, Predicasts, SciSearch.

Amount of online use.
The unweighted mean is actually 96 hours per year. The average time for an enquiry being 1/10 of an hour.

Now that's the gist of the 1982 Online Information Centre survey. Our own survey carried out between April and October 1983 was very different primarily because it was a face to face survey by my colleague Alex Murphy who travelled the length and breadth of Britain talking to librarians about their online interests and preferences. The purpose of his survey was to discover the extent if any of the demand that there might be for an online service offering the bibliographical data in our Catalogue of British Official Publications Not Published by HMSO. I will tell you more about this particular database, but now I want to fill out the bare bones of the statistics I have already given you. However to enable you to put our report in perspective Alex visited a total of 76 libraries including 17 university libraries, 13 polytechnics, and 13 public libraries.

Other categories included other educational libraries, science libraries, government libraries, special libraries and business libraries. What did we learn? Well, we learned that the majority of libraries preferred Dialog to any other host. 23 libraries volunteered this information, quite unprompted, in the course of conversation. The reason they gave varied from 'forgotten how to use any other host' from a Scottish College of Technology, to 'spread of files, easy command language, reliability from a large Midlands University Library. Five libraries preferred other hosts mainly because they were cheaper particularly Datastar; on the other hand the librarian of a major oil company said she used Dialog even when Datastar was cheaper because its searches are so good. There was no doubt that the majority of librarians did not like having to learn new search techniques and add extra pass words; a number of users talked about quick and dirty searches being all they did. It was significant that only in Science, Medicine and Engineering did they say that tney structured their searches and used thesauri. Amount of use of course varies widely from 'searches every minute of the day' in the library of a US oil company in London to two or three searches a month in a Scottish public library.

Combining the comments made to us with written comments from the '82 survey - we learn that the current proliferation of hosts and databases is actually confusing for many users who do not welcome it, who have difficulty in keeping up to date, who dislike the variety of search techniques they have to learn and who wish that there was a common command language to cover European and US hosts. It is strange to see such conservatism in such a new media but this is the reality. We had remarkably few comments about cost but there is a dislike of subscription charges and of dedicated terminals. Pay-as-you-use is seen as the only acceptable method of payment.

Some respondents suggested databases not at that time online that they would like to see online. 29 databases were suggested, including British Humanities Index, the Index to the London Times, Computer Journal Abstracts and data on companies from registrations in Companies House, and on yes our own Non-HMSO file.

So what is this file that we are now proposing to put online? Official publications from British government departments and agencies are traditionally published by HMSO - Her Majesty's Stationery Office. But over the last 20 years an increasing number of agencies have been using their own publications often in very small numbers for specialist audiences with no proper distribution or cataloguing.

This has created a chaotic bibliographical situation for librarians who began to realise that as much as 50% of all official publications were bypassing HMSO and the British National Bibliography and were thereby unknown and uncontrolled. We stepped into the breach _____ and in 1981 issued the first Catalogue of British Official Publications Not Published by HMSO covering the publications of nearly 400 agencies including every major government department, Treasury, Defence, Home Office and so on. We are sent copies of publications by all these agencies and we catalogue and index them and then microfilm the items themselves. The Catalogue is issued every two months and is cumulated at the end of each year. It now totals over 10,000 items a year compared to 8,000 HMSO titles per year so does indeed comprise 50% of official publications in the UK. The data is held in machine readable form and is an obvious candidate for online access.

Our survey revealed two things: first that the libraries already subscribing $260 a year to the hard copy Catalogue would not cancel if it was available online and second that there were many specialist and smaller libraries who could not justify the subscription price because of the narrowness of their interest but were eager to use it online.

The other file that British users want is HMSO's catalogue of its own publications - remember, these include important categories such as Parliamentary Papers, as well as many technical and trade publications. Now, what would be more logical than the linking of the HMSO and Non-HMSO databases so that they can be searched together as one Government Publications File and thus break down this division between HMSO and non-HMSO which most end-users regard as an absurd and incomprehensible barrier to finding the information they want?

It would also make sense to put up our Non-HMSO database with a host who already has a stable of public affairs databases. There is one - Scicon, a computer service company, part of British Petroleum. It has 2 databases which are particularly complementary to our official publications database. They are POLIS and Hansard.

POLIS is an index of Parliamentary information compiled by the staff of the House of Commons Library. Parliamentary questions, proceedings and Parliamentary papers are included while European Community, international and foreign official publications, press comments and other publications taken by the Library are included on a selective basis.

HANSARD is the full text of Hansard, the daily verbatim record published by HMSO of what is said in Debates in the House of Commons.

It appears to make sense for both HMSO and Non-HMSO bibliographic files to be available through this host.

But let's look at the dominant characteristics of both files: Taken together an annual increase of 18,000 publications, a lot of titles are spread quite thinly over a large number of subject areas. So would a researcher searching in one subject area confine himself to a search of such a database? The answer must be no. He will go to the specialist indepth databases on his subject and then add a further search of our wide ranging files to pick up references of significant publications which often do not get cited in any other bibliography.

This is what our 1983 survey so clearly told us. And Dialog is the host on which users want to conduct their searches.

I now know that the Non-HMSO database should be alongside indepth databases such as Chemical Abstracts, Aqualine and others. Alongside these it will get used, isolated from them it will not. HMSO will go online with Dialog later this year when they have finished converting the file into the MARC format and it is with great personal satisfaction that I can announce at this meeting, publicly, for the first time that we are signing an agreement with Dialog very shortly so that HMSO and non-HMSO databases can be made available from the same host, possibly as a combined database. I am sure that the two files will get a lot of use. The fact that nearly all the non-HMSO publications are available from us on microfiche at a very low unit cost will make it a very practical file to use.

I am now going to talk about other interesting home-grown databases in the UK across a complete spectrum of subject areas, and it is appropriate to start by mentioning 3 new databases that are part of Scicon's rapidly growing and very interesting staole. They are: Acompline and Urbaline from the Greater

London Council (the GLC) Research Library, and DHSS Data. Acompline provides indexing to all aspects of urban policies. First developed in 1973 it will contain 100,000 references by the end of this year. Urbaline launched on 4 June this year is the Daily Intelligence Bulletin and Weekly Technical Bulletin of the GLC with 83,000 news items to date from 1981.

Finally, DHSS Data is the database covering health services, social welfare and social issues produced by the Library of the Department of Health and Social Security. It was launched last Monday and 12,000 records will be added each year including coverage of US government publications and the publications of the World Health Organization.

Scicon has had some difficulty in meeting US demand because of their lack of representation in the US. However they are rewriting their user guide and are offering practice files at a reduced price so that you can both train and test yourself in the use of their databases.

I mentioned ASLIB, the parent of the Online Information Centre. The other main library association in the UK is The Library Association which publishes the LISA database, Library and Information Science Abstracts. Abstracts are prepared by the staff of the Library Association Library from 550 titles, mainly journals. LISA is available on 3 hosts: Dialog, Derwent-SDC and Samson Datanet.

I don't think my friends in the British Library would forgive me if I failed to mention BLAISE again, the British Library's online service which supplies the MARC files for cataloguers throughout the UK and abroad. BLAISE may be number two to Dialog but it is of very central importance to the majority of British libraries.

Other British databases held by BLAISE include AVMARC, bibliographic records of non-book materials, the British Education Index and Conference Proceedings.

I only know of two bibliographic databases in the field of art history, RILA and Art Bibliographies Modern but ArtQuest, a new online service from Art Sales Index Ltd is rather different since it is non-bibliographic and contains details of half a million oil paintings, watercolours and drawings sold at public auctions throughout the world since 1970. It is updated twice a month and is increasing by about 60,000 artworks a year.

It can be interrogated by index points such as nationality and period, and physical details such as medium-oil, or watercolour and height and width and whether signed. So it is a very useful database for searching for pictures about which you only have limited information. For example, you might ask: I have a watercolour by David Cox measuring 30" x 50', have any other watercolors by Cox of this size been sold recently and if so for how much? Or, 'What is the median and highest prices paid for works by abstract expressionists over the last four years?' ArtQuest is accessible through Pergamon Infoline.

One of the best known bibliographic database publishers in the UK is PIRA, the research association for the Paper and Board, Printing and Packaging Industries. There are five databases in all: International Packaging Abstracts, Printing Abstracts, and Paper and Board Abstracts. And two somewhat unrelated databases, Electronic Publishing Abstracts, which covers the world's literature since 1979, and Management and Marketing Abstracts which covers over 100 journals since 1976. All PIRA databases are available through Pergamon Infoline.

Moving on into other technical fields there is World Textiles. an online database produced by the Shirley Institute, a research and development institute established in Manchester the heart of the cotton industry 65 years ago, originally financed by a levy on the number of spindles in each cotton mill. The World Textiles Database is produced from the twice monthly literature survey periodical World Textile Abstracts, from 1970 onwards. By 1983, 110,000 items of literature were recorded. The file is available on Dialog and Pergamon Infoline

Weldasearch is the database service of the British Welding Institute and serves a truly international audience with subscribers in India, Brazil and Japan as well as the USA. Weldasearch comprises indexed abstracts from world welding literature from 1967. The file contains 70,000 items growing at a rate of 6,000 per annum. It is available on Dialog.

Laboratory Hazards is a useful database for researchers and laboratory staff compiled by the Royal Society of Chemistry. It covers reference books, journals, legislation and standards since July 1981 and is accessible through Pergamon Infoline and shortly through ESA and Datastar.

Chemical Hazards in Industry which started as a current awareness bulletin in January 1984 will be available online later this year. Another small but very valuable database.

The next group, in the Food Industry are four acronymically elegant files, FROSTI, FAST, FLAIRS, and FOMAD. Only the first exists the rest are 'about to be launched'. They are produced by the British Food Manufacturing Industries Research Association,

FROSTI publishes references in food science and technology taken from journals, foreign patents, conference proceedings books and other sources. The database holds 125,000 items from 1975 adding over 14,000 references a year. The database service is intended primarily for the association's members but will be accessible to US users through Telenet in the near future when a new subscription system for outsiders is set up.

The databases about to be launched are FAST which lists additives and suppliers (mainly in the UK), FLAIRS which provides food launch awareness in the retail sector (whatever that means) and FOMAD which is a source of food market information.

These are good examples of the many specialised databases in the UK that one does not often hear of.

One of the most interesting new databases is World Reporter created by the BBC and Datasolve, a computer service company .

World Reporter is an online news and current affairs information service, and is huge. It is only 2 years old and already contains over 100 million words of world wide news. Its information comes from 4 main sources: the BBC Summary of World Broadcasts from the BBC Monitoring Service which listens into and translates broadcasts in 40 different languages from 120 countries particularly the USSR, Eastern Europe, the Far East and Middle East, Africa and Latin America. The second source is BBC External Services News with stories by BBC reporters and correspondents in the UK and all over the world. It is particularly strong in Africa, the Middle East and Far East. Both these BBC sources date from January 1982.

Other sources are being added. These are the Economist, the weekly newspaper, back to 1982. Associated Press, the general news wire service, adding 55,000 words a day, and shortly the Guardian, one of the best British national newspapers and later this year the Washington Post.

Taking these sources together I think that World Reporter must be one of the most valuable general news information services in the world. US users access World Reporter through Datasolve's New York Office.

In the financial world Exshare, the financial database owned jointly by Extel and Reuters, has been getting a lot of attention in view of Reuters' successful London and New York share issue a few weeks ago. Exshare contains information on over 50,000 quoted securities on the world's leading stock markets including North America - but North America subscribers don't get North American securities everybody else does, presumably because since they are accessible through other North American services.

Exshare is supplied on magnetic tape or disk or diskette or by data transmission directly into the customer's computer. The service is daily, weekly or monthly.

There are two very valuable new databases from Inter Company Comparisons Ltd, an established information provider in the UK, called ICC Financial Data and ICC Directory of Companies.

ICC Financial Data provides comprehensive financial details taken from annual accounts submitted to Companies House, the national companies registration office on 60,000 private and public companies.

The ICC Directory of Companies updated weekly gives registration numbers and other brief details relating to annual accounts of 800,000 companies in the UK. It is a very quick way of identifying a company and discovering what documents have been filed by that company at Companies House.

To be able to get information on competitors, customers and suppliers online immediately is a very useful development indeed. I will use it and I know that many businessmen will be going to their public library for just this information. The databases are accessible on Dialog, via Prestel or by direct link.

There are of course many databases that are made available to the user by the physical transfer of tapes to the customers system rather than by online access.

For example, Lloyds Register of Shipping together with Lloyds, the London Insurance Corporation run various shipping information services including a New Construction File of ships on order or being built throughout the world, a file of all known sea going merchant ships of more than 100 tons, shipping movement services giving details of the movements of 32,000 merchant ships with 6,000 daily updates. There is a Shipowner and Parent Company File and a Casualty File reporting all serious casualties and total losses to ships of 100 tons and above since January 1978.

The data comes from every conceivable source including 1,700 Lloyds register surveyors throughout the world and a lady who occupies an apartment in a high rise block overlooking the harbour of Tampa, Florida. Lloyds also maintain a watch on the movements of certain ships for lawyers all over the world who will be interested in such ships because of ownership disputes, and unpaid bills and harbour dues.

Customers in the UK and US get their information on tape once every 3 months. Others take COM microfiche or ask Lloyds to answer specific questions for them. Lloyds will not have an online facility for at least 14 months.

One of the most prolific database publishers is the Commonwealth Agricultural Bureau with 30 databases. The Bureau acts as a clearing house for the collection and dissemination of agricultural information worldwide. The total number of records on file since January 1973 is over one million and is increasing at 150,000 per annum. The range of subjects is remarkable from Cotton and Tropical Fibres to Potatoes and Bee Keeping. The online files are searchable through Dialog, ESA or DIMDI (the German host).

Researchers in many universities create their own specialised databases and make them available to researchers elswhere. An interesting example in the field of ornithology is the FULMAR database created by scientists at the University of Aberdeen. FULMAR records data collected from observations of the Fulmar Petrel, a gull-like member of the Petrel family, that nests on the island of Eynhallow in the Orkney Islands, North of Scotland.

Observations have been accumulated since 1950 - an extraordinarily long period of continuous record keeping. Data collected about nest sites and the birds themselves enables survival rates and life expectancies to be assessed.

The University of Aberdeen computing centre also operates a database on experimental flocks of sheep and their ancestors, and a database on ante-natal care data and its results at Aberdeen Maternity Hospital. There is also the Wage Rounds Database which records details of all wage settlements of UK manual workers in nearly 200 negotiating groups since 1950.

These databases sound specialised but attract over 2,000 users even though they are not available through a commercial host. They could well become part of the Scottish Datacentre, a new idea, the intention for which is that there should be a collection of databases held at Edinburgh University which could be accessed online by users in Scotland and elsewhere. We have for example discussed the possibility of loading the non-HMSO publications file and our forthcoming machine readable subject catalogue of the 19th century parliamentary papers. Other databases could include Scottish census data, and boundary files for Scotland which already exist in digitised form from which maps can be drawn.

My guess is that the funding they need will be forthcoming and we will hear a lot more about the Scottish Data Centre.

I will end with mention of a new machine readable file that we are offering on tape and which may one day be accessible online. This is the Historical Names Database. Its core is our index of personal names in every contemporary British and Irish biographical dictionary published between 1840 and 1940. We estimate that the 360 titles many of which are multi-volume contain a staggering 6½ million entries on some 4 million people. We are microfilming the dictionaries and providing the index on COM or tape as part of the package.

However we will also include in the database all personal names in the National Union Catalog of Manuscripts Collection since we have just finished inputting these names for a hardcopy Index to Personal Names in the National Union Catalog of Manuscripts Collections 1959-1982; we will also include all personal names in the indexes to the National Inventories of Documentary Sources in the United States, United Kingdom and Ireland. The National Inventories reproduce,

on microfiche unpublished finding aids to archives and manuscript collections accompanied by indexes. For example in the National Inventory we include on microfiche the registers of collections in the Manuscript Division of the Library of Congress, and finding aids in the National Archives and the seven presidential libraries.

The indexes to these finding aids that we create contain the names of many thousands of people whose names may not be recorded anywhere else. It offers exciting possibilities through its sheer size and diversity and through being drawn from sources both sides of the Atlantic. An obscure emigrant from the UK or Ireland referred to in an American archives may turn out to be or to be related to the subject of a biography. It will be a huge treasure house for historians and genealogists or genies as I heard them referred to in Austin yesterday, both sides of the Atlantic.

I hope I have shown through all this that database creation in Britain is alive and well with lots of new things being developed. Future developments will be influenced by the attitudes of users and while it appears that there are no revolutionaries in our libraries, quite the contrary, it is those users whose views we must listen to if we are to create the databases that will successfully serve them in the future.